

Wavelets in Multi-Scale Time Series Analysis: An Application to Seismic Data



Stefania Corsaro , Pasquale Luigi De Angelis , Ugo Fiore ,
Zelda Marino , Francesca Perla , and Mariafortuna Pietroluongo 

Abstract Forecasting earthquakes is one of the most important problems in Earth science because of their devastating consequences. Current scientific studies related to earthquake forecasting focus on three key points: when the event will occur, where it will occur, and how large it will be. In this work we investigate the possibility to determine when the earthquake will take place.

We formulate the problem as a multiple change-point detection in the time series. In particular, we refer to the multi-scale formulation described in Fryzlewicz (Ann Stat 46(6B): 3390–3421, 2018). In that paper a bottom-up hierarchical structure is defined. At each stage, multiple neighbor regions which are recognized to correspond to locally constant underlying signal are merged. Due to their multi-scale structure, wavelets are suitable as basis functions, since the coefficients of the representation contain local information. The preprocessing stage involves the discrete unbalanced Haar transform, which is a wavelet decomposition of one-dimensional data with respect to an orthonormal Haar-like basis, where jumps in the basis vectors do not necessarily occur in the middle of their support.

The algorithm is tested on data from a well-characterized laboratory system described in Rouet-Leduc et al. (Geophys Res Lett 44(18): 9276–9282, 2017).

Keywords First keyword · Second keyword · Another keyword

Authors acknowledge the financial support provided by the Research grant of Università Parthenope, DR no. 953, november 28th, 2016.

S. Corsaro · P. L. De Angelis · U. Fiore (✉) · Z. Marino · F. Perla · M. Pietroluongo
Parthenope University, Napoli, Italy
e-mail: stefania.corsaro@uniparthenope.it; pasqualeLuigiDe.angelis@uniparthenope.it;
ugo.fiore@uniparthenope.it; zelda.marino@uniparthenope.it; francesca.perla@uniparthenope.it;
mariafortuna.pietroluongo@uniparthenope.it

1 Introduction

Seismic signals recorded on the field result from the interaction of the original source with the process of wave propagation. Experiments where seismic phenomena are induced in laboratory create (partially) controlled environments where the dynamics of earthquakes can be studied. Statistical learning methods are increasingly being used to isolate patterns in seismic signals that cannot be easily detected with traditional waveform analysis techniques [7]. Recently, a study on data originated from laboratory friction experiments [9] has investigated the possibility that natural earthquakes could be preceded by precursory signals, so that the detection and measurement of these signals could be used in forecasting.

When applying an analysis and forecasting model to very long signals, such as those related to seismic events, the hypotheses made in the model about the data-generating process may not necessarily be valid for the whole duration of the signal. Further, adaptation of the model parameters to data may become increasingly complex and time-consuming. In this perspective, transforming batches of the original signal into compact representations and observing the variation over time of such representations could be a valid solution. An evaluation of the potential of a transformation based on bottom-up data decomposition and change-point detection through wavelets has been carried out in this study. The next section briefly recapitulates some ideas about wavelet transforms as tools to estimate change points in piecewise-constant functions. Section 3 describes the data analyzed, and Sect. 4 discusses the experiments performed and their results.

2 Background

Wavelet thresholding estimators have received much attention in literature, since wavelet functions show some relevant properties. The key property of wavelets is referred to as “localization,” which allows one to obtain sparse representation of certain functions and operators in wavelet bases. For this reason, wavelet techniques can provide insight beyond other approaches in jump detection in high-frequency data. Traditional wavelet thresholding estimation proceeds as follows. Take the discrete wavelet transform of a dataset, set to 0 those coefficients that fall below a certain threshold, and then take the inverse wavelet transform of the thresholded coefficients. The definition of wavelet is a quite general one; thus many wavelet families can be built. They are classified according to certain properties such as orthogonality, amplitude of the support, smoothness, and the number of vanishing moments. Each of these properties is important for specific purposes; thus the choice of the wavelet basis is strongly application dependent. When the focus is data compression, smoothness, and a compact, narrow support is desirable: in this case, localization is improved, so that small coefficients are obtained in smooth regions of the approximated function. They can therefore be neglected, preserving information

about sub-domains in which the gradient has high values. Wavelet thresholding has indeed successfully been applied in several fields such as signal denoising, image analysis, and finance [1, 2, 4].

Using Haar wavelets one obtains piecewise-constant estimates. Piecewise-constant estimators are easy to interpret: jumps in the estimate can be viewed as relevant changes in the mean level of the data, whereas constant intervals represent periods in which the mean of the data does not significantly change. This feature makes them attractive in the field of earthquake forecast, in terms of time. In this case, one can formulate the problem as a multiple change-point detection in the time series of acoustic data. A posteriori detection of multiple change points, sometimes referred to as segmentation, can often serve as the useful first step in the exploratory analysis of data. Moreover, piecewise-constant estimates are cheap to store, because the number of jumps is typically significantly less than the size of the analyzed time series. This is relevant in our application, since a huge volume of data is to be taken into account. Nonlinear estimators exhibit superior theoretical and practical performance with respect to linear ones when the underlying function is spatially inhomogeneous. In [3] authors use piecewise-constant approximation to control the number of local extremes. On the other hand, a disadvantage of Haar thresholding is that, due to Haar wavelet construction, jumps always occur at dyadic locations, even if it is not justified by the data. In [6] authors introduced the unbalanced Haar (UH) wavelet basis, in which unlike traditional Haar wavelets, jumps in the basis functions do not necessarily occur in the middle of their support. Thus, they are potentially useful as building blocks for piecewise-constant estimators that avoid the restriction of jumps occurring at dyadic locations. These wavelets enjoy the desirable properties of traditional wavelets, such as a multiresolution structure and an associated fast transform algorithm.

Our estimation procedure can be summarized as follows. We first take a transform of the data with respect to an UH basis. We then threshold the coefficients and take the inverse transform.

3 Data

The data used to test our model comes from a laboratory earthquake experiment described in [9]:

- The input is a chunk of 0.0375 s of seismic data (ordered in time), which is recorded at 4 MHz, hence 150,000 data points, and the output is time remaining until the following lab earthquake, in seconds;
- the seismic data is recorded using a piezoceramic sensor, which outputs a voltage upon deformation by incoming seismic waves. The seismic data of the input is this recorded voltage, in integers;
- seismic data include both a training set and a testing set, which come from the same experiment. There is no overlap between the training and testing sets, which

Table 1 Summary of the 17 blocks the data was found to be divided in. For each block, the start and end times are given, together with the number of measurements

Block no.	Time first	Time last	No. of readings
1	1.4691	0.0008	5656574
2	11.5408	0.0007	44429304
3	14.1806	0.0008	54591478
4	8.8567	0.0011	34095097
5	12.6940	0.0006	48869367
6	8.0555	0.0011	31010810
7	7.0590	0.0005	27176955
8	16.1074	0.0007	62009332
9	7.9057	0.0002	30437370
10	9.6372	0.0005	37101561
11	11.4264	0.0002	43991032
12	11.0242	0.0001	42442743
13	8.8281	0.0001	33988602
14	8.5660	0.0009	32976890
15	14.7518	0.0006	56791029
16	9.4595	0.0006	36417529
17	11.6186	9.7598	7159806

are contiguous in time. However, since no ground truth is available for the testing set, in this work n -fold cross-validation has been performed on the training set only;

- time to failure is based on a measure of fault strength (shear stress, not part of the published data). When a labquake occurs, this stress drops unambiguously;
- data is recorded in *bins* of 4096 samples. Within those bins seismic data is recorded at 4 MHz, but there is a 12-microsecond gap between each bin, an artifact of the recording device.

In addition, additional structure was found by examining the seismic data. The training set was found to be subdivided into 17 *blocks* of varying length, separated by different time gaps (see Table 1 for details).

To gain some insights about the data, an initial step involves computing and visualizing the autocorrelation. Figures 1 and 2 show, respectively, the autocorrelation and the partial autocorrelation averaged over all the bins of for the first data block. The charts for the subsequent blocks do not differ substantially and were not reported.

Recall that the autocorrelation is the correlation between y_t and y_{t-k} for different values of the lag k , while the partial autocorrelation gives the same correlation as above after the effects of the lags $1, 2, \dots, k-1$ have been removed. Autocorrelations have been averaged over all the bins to smooth out values which may be due to particular situations in each single 4096-measurement bin.

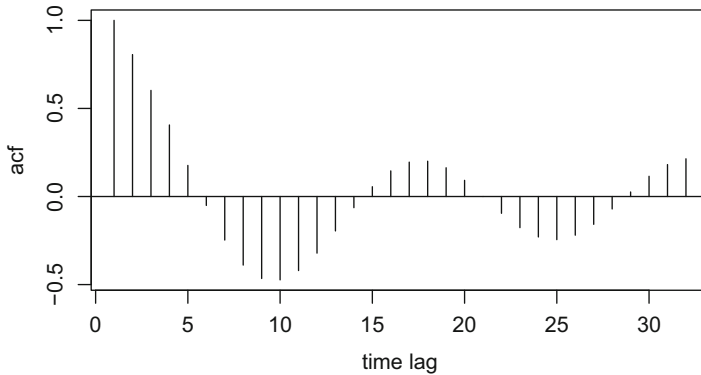


Fig. 1 Autocorrelation, computed for each of the 4096-readings bins of the first block of contiguous measurements, and then averaged

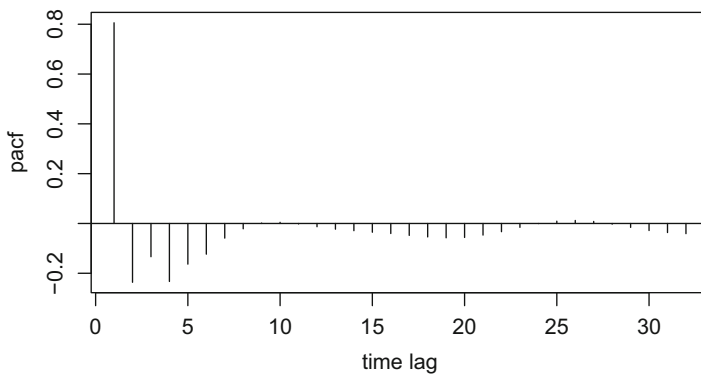


Fig. 2 Partial autocorrelation, computed for each of the 4096-readings bins of the first block of contiguous measurements, and then averaged

4 Experiments

The sheer size of the data would have had an adverse effect on the performance of training machine learning models. In addition, since data are recorded over a relatively long time with respect to the fine granularity of measurements, using a “flat” approach where each individual sample is taken separately did not look very attractive. In a hierarchical perspective, instead, if a way is found to condense each bin of readings in a representation in a small-dimension space, the evolution over time of this representation can be studied more easily.

The coefficient of a fitted AR(1) model – an autoregressive model of order one – was one of the features computed starting from a data bin. In fact, the damped sinusoidal shape of the autocorrelation, together with the presence of a spike at lag one in the partial autocorrelation, suggests [8] that fitting an AR(1) model to the data may be appropriate.

Since the transformation being sought should in some way capture the “energy” content of the observed signal, it is intuitive to think at entropy as a measure of uncertainty. The Shannon entropy of a discrete random variable Y is the expectation of the information content:

$$H(Y) = E_Y [-\log \Pr(Y)]$$

and, given a sample, it can be estimated from the observed counts. The entropy package of **R** has been used in the experiments. Note that, in all experiments, entropy was measured in bits.

The third transformation that has been used in the tests is the number of change points in the piecewise-constant mean of the noisy input vector, as described in Sect. 2. The efficient method implemented in the **R** package **breakfast** to estimate the number of change points was a critical factor in allowing the use of this technique, since computation times were reduced substantially [5].

Before going into further analysis, an interesting question that arises is at which scale the aggregation is to be performed. While the transformations can be applied to individual data bins, they could as well operate on sequences of contiguous bins (*windows*). Larger windows would tend to capture long-term effects, smoothing out fluctuations, whereas smaller windows would enable a more faithful description of short-lived variations. A preliminary calibration experiment was thus performed to select an appropriate window size. Out-of-sample correlation has been computed, after transforming the data in block number 6 (for training) and block number 7 (for testing) in different ways and for varying window sizes. Table 2 shows the results.

Entropy is seen to be the worst performer, while the number of change points obtains the best results, and the coefficient of an AR(1) model scores not too far. Moreover, a growth trend in correlation can be observed for all transformations as the window size is increased, suggesting that the accumulation of tension in the fault is a gradual process. Further experiments performed on the NCP transformation only for larger window sizes yielded correlation values as high as 0.881 for a windows size of 128 bins and even 0.945 for a window size of 256 bins. However, having the window size not exceed 32 bins – for a total of 131072 measurements – seemed to be appropriate, also in consideration that the number of 150,000 readings is used and mentioned often in [9]. A window size of 32 bins was therefore selected for the subsequent experiment.

Table 2 Out-of-sample correlation in calibration experiments for different transformations (ENT, entropy; AR1, coefficient of an AR(1) model; NCP, number of change points)

Window size (bins)	ENT	AR1	NCP
32	0.340	0.684	0.783
16	0.262	0.604	0.737
8	0.232	0.552	0.676
4	0.184	0.502	0.603
2	0.151	0.456	0.526
1	0.136	0.414	0.461

Table 3 Out-of-sample RMSE and correlation in 17-fold cross-validation experiments: training set is block i ; test set is all other blocks, $i = 1, \dots, 17$

	NCP	AR1
Mean RMSE	3.451	3.703
Mean correlation	0.603	0.501

All of the blocks in training data were used for a 17-fold cross-validation experiment. Each block was used to train a simple linear model from scratch, and all other blocks were used as testing data to verify the predictions. A linear regression model was purposely chosen as a very simple tool that would clearly expose the performance of the transformations being compared. The metrics selected to evaluate performance were the RMSE (root-mean-square error) and the correlation between the predicted data and the actual data.

Table 3 shows the results of the 17-fold cross-validation experiment. Note that n -fold cross-validation experiments are usually performed with n equal to 5 (or, less often, 10), implying a ratio between training and testing data of 1/5 (or 1/10). Here, the partitioning into 17 blocks means that the training-to-testing ratio is 1/15.84, taking into account the different sizes of the blocks.

The NCP transformation was found to outperform AR1. Finally, it was observed that using both methods in conjunction, i.e., fitting a linear model on both features, did not produce substantial improvements.

References

1. Chang, S.G., Vetterli, M.: Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing* **9**(9), 1532–1546 (2000). <https://doi.org/10.1109/83.862633>
2. Corsaro, S., D. Marazzina, D., Marino, Z.: A parallel wavelet-based pricing procedure for Asian options. *Quantitative Finance* **15**(1), 101–113 (2015). <https://doi.org/10.1080/14697688.2014.935465>
3. Davies, P.L., Kovac, A.: Local extremes, runs, strings and multiresolution. *The Annals of Statistics* **29**(1), 1–65 (02 2001). <https://doi.org/10.1214/aos/996986501>
4. Donoho, D.L., Johnstone, I.M.: Threshold selection for wavelet shrinkage of noisy data. In: *Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. vol. 1, pp. A24–A25 vol.1 (1994). <https://doi.org/10.1109/IEMBS.1994.412133>
5. Fryzlewicz, P.: Tail-greedy bottom-up data decompositions and fast multiple change-point detection. *The Annals of Statistics* **46**(6B), 3390–3421 (12 2018). <https://doi.org/10.1214/17-AOS1662>
6. Girardi, M., Sweldens, W.: A new class of unbalanced haar wavelets that form an unconditional basis for l_p on general measure spaces. *Journal of Fourier Analysis and Applications* **3**(4), 457–474 (Jul 1997). <https://doi.org/10.1007/BF02649107>
7. Holtzman, B.K., Paté, A., Paisley, J., Waldhauser, F., Repetto, D.: Machine learning reveals cyclic changes in seismic source spectra in geysers geothermal field. *Science advances* **4**(5) (2018)

8. Hyndman, R.J., Athanasopoulos, G.: Forecasting: principles and practice. OTexts (2018)
9. Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C.J., Johnson, P.A.: Machine learning predicts laboratory earthquakes. *Geophysical Research Letters* **44**(18), 9276–9282 (2017). <https://doi.org/10.1002/2017GL074677>, <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL074677>